VITON-GT: An Image-based Virtual Try-On Model with Geometric Transformations

Matteo Fincato\textsuperscript{1}, Federico Landi\textsuperscript{1}, Marcella Cornia\textsuperscript{1}, Fabio Cesari\textsuperscript{2}, Rita Cucchiara\textsuperscript{1}

\textsuperscript{1}University of Modena and Reggio Emilia, \textsuperscript{2}YOOX NET-A-PORTER GROUP

Email: \textsuperscript{1}\{name.surname\}@unimore.it, \textsuperscript{2}\{name.surname\}@ynap.com
• We propose a novel architecture for image-based 2D single-pose virtual try-on called **VITON-GT** (VIrual Try-ON with Geometric Transformations).

• The proposed model includes two different components: a two-stage geometric transformation module and a transformation-guided try-on module.
We employ two different geometric transformations, namely affine and thin-plate spline, to warp the in-shop image \( c \) of a particular garment.

- We compute the parameters \( \theta_1 \) for the affine transformation \( T_1 \).
- We perform regression to predict the parameters \( \theta_2 \) for the TPS transformation \( T_2 \).
- Final loss:
  \[
  L_{GT} = \lambda_1 L_{T_1} + \lambda_2 L_{T_2}
  \]
We generate an output image $\hat{I}$ representing the reference person wearing $c$ by employing a U-Net architecture.

We apply the previous learned spatial transformations in the clothes branch.

To guide the generation of the final image we implement three different losses: a $L_1$ loss, a perceptual loss and an adversarial loss.

$$L_{TON} = \rho_1 L_{ton} + \rho_2 L_{vgg} + \rho_3 L_{adv}$$
### Warping Results

<table>
<thead>
<tr>
<th>Try-on Garment</th>
<th>Target Model</th>
<th>CP-VTON [1]</th>
<th>VITON-GT</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Try-on Garment" /></td>
<td><img src="image2" alt="Target Model" /></td>
<td><img src="image3" alt="CP-VTON" /></td>
<td><img src="image4" alt="VITON-GT" /></td>
</tr>
<tr>
<td><img src="image5" alt="Try-on Garment" /></td>
<td><img src="image6" alt="Target Model" /></td>
<td><img src="image7" alt="CP-VTON" /></td>
<td><img src="image8" alt="VITON-GT" /></td>
</tr>
<tr>
<td><img src="image9" alt="Try-on Garment" /></td>
<td><img src="image10" alt="Target Model" /></td>
<td><img src="image11" alt="CP-VTON" /></td>
<td><img src="image12" alt="VITON-GT" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>FID</th>
<th>KID</th>
<th>IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-VTON [1] (TPS only)</td>
<td>101.12</td>
<td>6.80±0.67</td>
<td>3.31±0.35</td>
</tr>
<tr>
<td>VITON-GT (Affine + TPS)</td>
<td>59.53</td>
<td>3.27±0.48</td>
<td>3.40±0.22</td>
</tr>
</tbody>
</table>

- Our model outperforms CP-VTON [1] on all the evaluation metrics.
- The affine transformation helps the TPS generating better warped clothes.
- Reduced artifacts and distortions.

---

Try-On Results

- Reduced distortions while maintaining textures.
- Increased realism of the final images.
- Preserving details and characteristics of the original clothes.

Results on Out-of-Domain Clothes

Try-on Garment  Target Model  CP-VTON [1]  VITON-GT

Try-on Garment  Target Model  CP-VTON [1]  VITON-GT

Try-on Garment  Target Model  CP-VTON [1]  VITON-GT

<table>
<thead>
<tr>
<th>Model</th>
<th>Short-Sleeve T-Shirts</th>
<th>Long-Sleeve T-Shirts</th>
<th>Sleeveless T-Shirts</th>
<th>Shirts</th>
<th>Sweatshirts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FID</td>
<td>KID</td>
<td>IS</td>
<td>FID</td>
<td>KID</td>
</tr>
<tr>
<td>CP-VTON [1]</td>
<td>23.81</td>
<td>0.86±0.16</td>
<td>2.41±0.21</td>
<td>31.92</td>
<td>1.85±0.33</td>
</tr>
<tr>
<td>VITON-GT (no FT, no Adv. Loss)</td>
<td>22.11</td>
<td>0.76±0.16</td>
<td>2.54±0.12</td>
<td>23.74</td>
<td>0.89±0.22</td>
</tr>
<tr>
<td>VITON-GT (no Adv. Loss)</td>
<td>20.95</td>
<td>0.61±0.16</td>
<td>2.63±0.17</td>
<td>20.02</td>
<td>0.62±0.16</td>
</tr>
<tr>
<td>VITON-GT</td>
<td>20.73</td>
<td>0.57±0.15</td>
<td>2.65±0.14</td>
<td>20.83</td>
<td>0.64±0.17</td>
</tr>
</tbody>
</table>

• We have presented VITON-GT, a new image-based virtual try-on model that integrates multiple geometric transformations of the input clothes during the generation of the try-on result.

• Through extensive experiments on two different datasets, we have demonstrated the effectiveness of our solution w.r.t. previously proposed methods.
Thank you!

Matteo Fincato$^1$, Federico Landi$^1$, Marcella Cornia$^1$, Fabio Cesari$^2$, Rita Cucchiara$^1$

$^1$University of Modena and Reggio Emilia, $^2$YOOX NET-A-PORTER GROUP

Email: $^1$\{name.surname\}@unimore.it, $^2$\{name.surname\}@ynap.com